

**University of Colorado, Boulder**  
**CU Scholar**

---

Undergraduate Honors Theses

Honors Program

---

Spring 2013

# Response Window Length in the Weapon Identification Task: How Executive Function Ability Modulates Implicit Racial Bias

Anna Gitarts

*University of Colorado Boulder*

Follow this and additional works at: [http://scholar.colorado.edu/honr\\_theses](http://scholar.colorado.edu/honr_theses)

---

## Recommended Citation

Gitarts, Anna, "Response Window Length in the Weapon Identification Task: How Executive Function Ability Modulates Implicit Racial Bias" (2013). *Undergraduate Honors Theses*. Paper 359.

This Thesis is brought to you for free and open access by Honors Program at CU Scholar. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of CU Scholar. For more information, please contact [cuscholaradmin@colorado.edu](mailto:cuscholaradmin@colorado.edu).

Running Head: EF ABILITY AND IMPLICIT RACE BIAS

Response Window Length in the Weapon Identification Task:  
How Executive Function Ability Modulates Implicit Racial Bias

Anna Gitarts  
Department of Psychology and Neuroscience  
University of Colorado Boulder

Thesis Advisor: Dr. Tiffany A. Ito  
Department of Psychology and Neuroscience

Defense Committee:  
Dr. Richard K. Olson, Honors Representative, Department of Psychology and Neuroscience  
Dr. Hillary Potter, Department of Sociology

April 2, 2013

## ABSTRACT

The purpose of this study was to determine how executive function (EF) ability modulates stereotype-derived responses on a specific implicit bias task at different response window lengths. Each participant completed two EF tasks and one version of the weapon identification task (WIT). The EF tasks were used to measure participants' abilities to control their responses and the WIT was used as a measure of implicit racial bias. It was found that participants made more stereotypical misidentifications at shorter response windows and responded faster to stereotype-congruent trials at longer response windows, both of which directly replicate previous research. But, EF ability could only be used to predict bias in reaction times. This finding does not replicate previous research findings for stereotypical misidentifications, despite similarities in experimental design between this study and previous research. Because of the low number of participants per condition and the lack of consistent patterns of results, there are significant limitations in the conclusions that can be drawn from this study. Reasons for why the results were sporadically significant are discussed.

## ACKNOWLEDGEMENTS

This thesis would not have been possible without the mentorship and assistance of my direct supervisor Dr. Chris Loersch and my thesis advisor Dr. Tiffany Ito. They have made it easier to understand an immensely complicated area of research that I could not have grasped without their help. I would also like to thank Chris Lindholm for his support, edits, and time spent listening to me talk endlessly about this project. Finally, I would like to thank Laxmi Lalchandani for assisting with data collection.

## Response Window Length in the Weapon Identification Task:

### How Executive Function Ability Modulates Implicit Racial Bias

Automatic and controlled cognitive processes have been the subjects of a large, and ever-increasing, body of research. Historically, controlled processes have been defined as processes within conscious control that can be changed and require significant effort and attention. On the other hand, automatic processes have been defined as processes outside of conscious control that cannot be purposefully altered. The difference between automatic and controlled processes was demonstrated by Devine's (1989) work on stereotype activation. In this research, she showed that when participants were subliminally exposed to stereotype-related words associated with Blacks (versus neutral words unrelated to race), they were more likely to rate a race-unspecified character's actions in a narrative as hostile, regardless of the participants' levels of prejudice. In contrast, when participants were directly asked to list their beliefs about Blacks, highly prejudiced participants were more likely to include character traits such as hostility and aggressiveness compared to less-prejudiced participants. Together, these two studies demonstrate that stereotypes are automatically activated regardless of people's conscious beliefs, and that people who are motivated to control their prejudice are only able to do so when they are aware that the stereotypes have been activated.

To date, much of the research on automatic and controlled processes has focused on prejudice and stereotypes, as in the Devine (1989) work. Originally, researchers used two types of explicit tasks to measure stereotyping and prejudice: direct and indirect tasks. Direct explicit tasks directly ask participants about their racial attitudes (these are tasks such as surveys and questionnaires). Indirect explicit tasks indirectly measure attitudes by obscuring the purpose of

the task to decrease the likelihood that participants will modify their responses to seem less biased. An example of an indirect task is the Funding task used in the current experiment, in which participants are asked to rank different student organizations by their relative importance, including organizations related to minority groups. Though the task does not directly ask for opinions on minorities, attitudes toward minority groups can be inferred from the rankings given to these organizations over other ones. It is still possible for participants to rank minority-related organizations as more important than other organizations in order to seem less biased, but this possibility is less likely when race is not made salient in the task instructions.

Eventually, researchers found that explicit tasks often fail to accurately predict prejudiced actions, potentially because participants may be motivated to seem less biased (Devine, 1989). Therefore, there was a call to develop tasks that measure unconscious processes of stereotype activation separately from conscious ones, and implicit tasks were developed (Greenwald & Banaji, 1995; Payne & Gawronski, 2010). Unlike direct explicit tasks, implicit tasks can be used to indirectly measure attitudes by obscuring the purpose of the task and infer attitudes or stereotypes from participants' responses, which is similar to how indirect explicit tasks measure attitudes. But unlike indirect tasks, implicit tasks are used to infer bias from the pattern of responses participants give on a task that is unrelated to racial attitudes (such as a simple categorization task). By obscuring the purpose of the task and not asking participants to consciously give their attitudes, it was thought that participants would not be able to respond in a "socially desirable manner" and that these tasks were only a measure of automatic activation of stereotypes.

However, research has shown that scores on implicit bias tasks are malleable. In this context, malleability means that the participants' responses on implicit bias tasks differ

significantly when the researcher changes the task. This suggests that either different implicit attitudes are activated under different conditions, or that participants are actually able to control the amount of bias they express on the tasks. If participants can control the amount of bias they express, then they can still respond in a socially desirable manner, despite the fact that they are not explicitly stating or being asked for their attitudes. Two types of task manipulations have commonly been used in previous research to demonstrate malleability of implicit task scores.

The first are contextual changes, which alter what stereotypes are activated. Blair (2002) details many examples of contextual changes to tasks, including work by Wittenbrink, Judd, and Park (2001), in which participants were shown either a video of a Black family smiling at a barbeque or a video of angry Black gang members. After being shown the video, participants completed a specific implicit bias task, the implicit association test (IAT). Participants who had been shown the happy family had lower racial bias scores than participants who had been shown the gang members, which indicates that different stereotypes can be activated depending on what beliefs have recently been made salient about a particular social or racial group.

The second type of modification is operational (i.e. procedural), which changes how easily a person can consciously analyze the situation and make a decision. For example, Payne, Lambert, and Jacoby (2002) showed that changing the response window length affected the number of stereotype-congruent errors participants made on implicit tasks: shorter response windows elicited more errors. If participants are able to control their responses and consciously reduce the amount of bias they express when the response window is longer, but show more bias when the response window is shorter, then implicit tasks cannot only be measuring automatically activated attitudes. This research suggests that implicit tasks measure both controlled and automatic processes. Like the aforementioned research, the current study focuses on operational

changes – specifically, response window changes – and how these changes affect scores on implicit bias tasks.

The research on malleability has indicated that implicit bias task scores have both controlled and automatic components for every trial. Rather than considering automatic and controlled processes to be completely distinct, there has been a paradigm shift in the field, and now these processes are considered to be on a continuum for all responses. Thus, multiple models for determining the automatic and controlled components of scores have been created. One of the most commonly used is a dual-process model, the process dissociation procedure (PDP), formulated by Jacoby (1991). The PDP model uses separate equations to estimate the controlled and automatic components of implicit task responses – the automatic estimate is assumed to directly reflect automatic stereotyping and the controlled component is assumed to directly reflect cognitive control. The controlled component indicates how often participants respond correctly to trials because they were following task instructions. The automatic component, on the other hand, indicates how often participant responses were guided by automatic stereotype activation when control failed (see Jacoby, 1991 and Payne et al., 2002 for more details on calculating PDP estimates and the theory behind PDP). Using these PDP estimates, Payne et al. (2002) showed that at shorter response deadlines participants were less able to control their responses (lower PDP control estimates), but that PDP automatic estimates were unchanged by the response deadline manipulation. This result demonstrates that responses on implicit tasks are driven by both controlled and automatic processes, but that the magnitude of the controlled component depends on the response window length.

Because the controlled component of the PDP is assumed to directly indicate cognitive control, some newer research has focused on how executive function ability modulates scores on

these implicit tasks. Executive function (EF) ability (i.e. executive control) is a fixed measure of how well people are able to control their responses. As shown in the Payne et al. (2002) study, how well participants were able to control their responses was influenced by the response window length. Unfortunately, since the PDP only gives indirect estimates of cognitive control, a verification that the controlled component does correlate with executive function ability was needed. In the Payne (2005) work, PDP control estimates were compared to EF ability by having participants do an EF task, the Antisaccade task, in addition to the implicit bias tasks. Antisaccade measures participants' abilities to inhibit their automatic responses, which indicates how easily they are able to follow task instructions. The Payne (2005) study confirmed that EF ability and the PDP control component correlate with one another, thus solidifying the connection between EF and operational malleability of implicit bias scores.

A related issue in recent implicit bias research is how to reduce the influence of executive control on implicit bias scores. For example, work by Siegel, Dougherty, and Huber (2012) looked at separating the influence of executive control from bias on the IAT (a specific implicit bias task) by inducing anxiety in the participants and using explicit tasks as a measure of racial bias separate from the IAT. They proposed that inducing anxiety decreases participants' abilities to use executive control by taking up some of the attentional resources necessary for following task instructions. The results indicated that decreasing executive control ability increased the amount of racial bias participants expressed on the IAT. From this manipulation, Siegel et al. (2012) concluded that executive control is not only correlated with IAT scores, but that it directly affects performance on the IAT.



### The Current Study

The current study looks at how the effects of executive function ability on racial bias change under different response window lengths on the weapon identification task (WIT; Payne, 2001). Though the Payne et al. (2002) study compared participant PDP control estimates to their scores on the WIT, actual EF tasks were not used to measure EF ability. Similarly, the Payne (2005) study used Antisaccade as an EF task, but did not look at how biases at different response window lengths were affected by EF ability. The current study uses two different EF tasks, Antisaccade and Spatial Span, to determine EF ability.

Because implicit bias task scores have been shown to be correlated with EF ability (Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010; Payne, 2005; Payne & Gawronski, 2010; Siegel et al., 2012), it was hypothesized that participants with higher scores on EF tasks, indicating a stronger ability to control their responses and follow task instructions, would generally show lower levels of racial bias on the WIT (racial bias is defined as greater accuracy on stereotype-congruent trials than on stereotype-incongruent trials).

In addition, decreasing the response window length has been demonstrated to inhibit use of EF (Payne, 2001; Payne et. al, 2002). Therefore, it was hypothesized that participants would generally have higher racial bias on the WIT when there was a shorter response window (400 ms) than when there was a longer one (3,000 ms), because there would be too little time to use executive control when the deadline was shorter. This main effect of response window was also expected to interact with EF ability. At longer response windows, participants with low and high EF ability were expected to have similar accuracy bias scores, because there would be enough time for all participants to use executive control. At shorter response windows, participants with low EF ability were expected to make more stereotype-congruent errors than participants with

high EF. Even though all participants would have little time to process the different trial types, participants with high EF would be able to control their responses better, as demonstrated by previous research (Payne, 2005). This means that participants with low EF would have higher accuracy bias scores than participants with high EF at these shorter response windows.

In summary, the main effect of EF task scores on accuracy bias on the WIT and the main effect of response window on accuracy bias on the WIT were expected to be significant. The interaction of EF ability X response window was also expected to be significant, indicating that the effect of response window length would be stronger for participants with low EF ability.

The predictions that have been discussed so far have defined racial bias only in terms of accuracy, but racial bias can also be considered in terms of how quickly participants respond to stereotype-congruent trials versus stereotype-incongruent trials. If stereotypes are influencing the speed of participants' responses, then there will be a reaction time bias (i.e. faster reactions on stereotype-congruent trials than on stereotype-incongruent trials). Because of this, the effects of EF scores and response window lengths for WIT reaction time bias were also evaluated. As previous research has demonstrated (Payne, 2001), when participants have sufficient time to use executive control (i.e. when the response window is longer), they make fewer mistakes, and instead respond faster to stereotype-congruent trials over stereotype-incongruent ones. This means that when participants have less time to respond, they are more likely to have larger accuracy biases. As the response window length increases, the accuracy bias will decrease and the reaction time bias will increase. Therefore, it was predicted that the main effect of response window on reaction time bias would be significant. It was unclear how a person's EF ability would be related to their RT bias, because it is possible that participants with higher or lower EF ability could react impulsively to stereotypical information. Therefore, no explicit predictions

were made for the main effect of EF ability on the RT bias or for the interaction of EF ability X response window length.

Although not of primary interest, the WIT biases were also compared to first-person shooter task (FPST) scores and scores on different explicit measures of racial bias. Including another implicit bias task (the FPST) and various direct and indirect explicit measures allowed an examination of how manipulating response window length changes the ability of the WIT to predict responses on other implicit tasks as well as on explicit tasks.

## Method

### *Participants*

One hundred and fifteen undergraduates at the University of Colorado completed the experiment for partial course credit. One participant was excluded from all analyses, because he pressed the same button for all WIT trials, regardless of the prime and target. Another participant was excluded from the EF analyses, because she was unable to complete the entire Antisaccade task (this individual was not excluded from the other analyses). All participants identified themselves as White.

### *Materials and Procedures*

All participants signed a consent form before the experiment. The consent form informed participants that the purpose of the study was to investigate “how cognitive control processes are related to various forms of computerized and paper-and-pencil judgment and decision tasks.” Participants received a debriefing sheet after completion of the experiment and were thanked for their time. Participants completed the following tasks in the order listed below.

The first two tasks were indirect explicit tasks:

*Funding Task:* Participants were given a letter informing them that the university needed to reduce funding for some of its student organizations and wanted input from the students. The letter asked them to rank a list of 29 organizations from 1 to 29 for continued funding, with 1 being most important to fund and 29 being least important to fund.

The list included organizations that involve different ethnic and racial groups (such as the African-American Student Association and the Argentine Tango Club) as well as other groups on campus (such as the Intervarsity Christian Fellowship and the Chess Club). In order to measure racial bias toward African Americans, scores on this task were calculated by averaging the ranks given to the Prospective Freshman Minority Visit Program, the Minority Arts and Sciences Program, the Summer Minority Access to Research Training Program, and the African-American Student Association. Participants who put these four organizations near the end of their lists were given higher bias scores.

*Seating Task:* In this task, participants were shown pictures of different places and were asked to pick which seat they would choose to sit in if they entered the room. One or more seats in the pictures were “occupied” with faces of people with different genders and races, and the participants were given the option of choosing one of four different seats. The seats the participants were able to choose varied in distance from the people already “seated” in the picture (see Figure 1 for an example scene). Participants were told that the purpose of the task was to find out how people use public spaces. Participants who chose to sit further from African Americans than from Caucasians were given higher racial bias scores.

Next, participants completed two commonly used executive function tasks (e.g. Friedman et al., 2008; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Shah & Miyake, 1996).

Participants were not told the purpose of these tasks:

*Antisaccade Task:* Participants were instructed to look at a fixation point in the center of the computer screen (a “+” symbol) until they saw a flashing cue that would appear on either the left or right side of the screen (the fixation point would disappear at this time). The cue would flash for 175 ms, 200 ms, or 225 ms, depending on the block. Then a digit, 1 through 9, would be displayed. After 150 ms, this target was masked by a shaded box. Participants were instructed to say the number they saw out loud, and the numbers were recorded by the experimenter on the keyboard. If the participants had not seen the number, they were instructed to tell the experimenter their “best guess.” Once the number was entered, the fixation point would reappear in the center of the screen and the next trial would begin.

In the first set of 12 practice trials and a block of 20 actual trials, the number appeared on the same side as the cue (prosaccade trials). In the next set of 12 practice trials and three blocks of 29 trials each, the number appeared on the opposite side of the cue (antisaccade trials). For the three antisaccade blocks, the number appeared more quickly with each subsequent block, such that it became increasingly more difficult to see the number after looking at the cue (the cue-to-target interval decreased from 225 ms to 200 ms to 175 ms). Participants were informed prior to each block that the number would appear more quickly.

This task measures inhibition ability, which is a type of executive function. Because participants are unable to look at both the cue and the number during the antisaccade trials, they must inhibit their reflexive (automatic) response to look at new stimuli in their visual fields (i.e. the flashing cue) in order to be able to see the number. Scores were calculated by looking at the

number of correct responses that were given on the antisaccade trials. A higher score indicates better inhibition ability. The reliability of this task has been shown to be .80 in previous research (Ito et al., 2013).

*Spatial Span:* Participants were first shown a capital letter (“P,” “F,” or “R”) that was rotated and either in its normal orientation (e.g. P) or mirrored across its vertical plane (e.g. P). Participants were instructed to tell the experimenter out loud whether the letter was “normal” or “mirrored” once the letter appeared on the screen. When the experimenter typed in “n” or “m,” respectively, on the keyboard, an arrow pointing in one of eight directions would appear on the screen for 750 ms and then another letter was shown. This sequence continued for a total of two to five iterations.

At the end of each trial, a pink “???” symbol would appear on the screen. Participants were instructed to remember the sequence and direction of the arrows they had seen and write them down on a piece of paper when they saw the “???” symbol. The paper had a box for each trial with lines to represent the possible arrow directions. Participants were instructed to write the sequence number next to each line that corresponded to the direction the arrows had flashed on the screen. Once the participants had written down all of the arrows they remembered, the experimenter would start the next trial. Participants were instructed to skip writing down any arrows they did not remember, because incorrect arrows (direction or sequence) would be marked as incorrect. Participants did not have a time limit for responding to the letters or for writing down the arrows.

This task measures working memory capacity (remembering the direction and sequence of the arrows) with distractors (the letters), which is a measure of executive function ability. Scores were calculated by determining the percentage of correct responses out of the total

possible. Higher scores represent better working memory capacity. The reliability of this task has been shown to be .91 in previous research (Shah & Miyake, 1996).

Following the EF tasks, participants completed two commonly used implicit bias tasks (e.g. Correll, Park, Judd, & Wittenbrink, 2002, 2007; Payne, 2001, 2005):

*Weapon Identification Task (WIT)*: Participants were instructed that the purpose of this task was to identify objects as guns or tools. First, participants were shown a black and white square with a scrambled pattern (mimicking static on a television screen) for 500 ms. Then, they were shown a picture of either a Black or a White male's face (200 ms), followed by a picture of a gun or a tool (200 ms). Participants were told that the picture of the face was their cue that an object was going to appear. Finally, the picture of the object was masked by another static square (presented for 300 ms). Participants were instructed to press a button on a button box labeled "gun" or "tool" to identify the picture of the object they had seen. Once participants pressed either button, the next trial would begin. Each trial also had a set response deadline (length depended on the condition, see below) in which the participant needed to press either button or else a "TOO SLOW!" message flashed on the screen for 500 ms and the next trial would begin after a 1,000 ms inter-trial interval.

Participants were given 10 practice trials, followed by one block of 128 actual trials. Before the actual trials, participants were instructed to respond "as QUICKLY AS POSSIBLE while also trying to be accurate."

Each participant completed one version of this task with a specific response window length: 400 ms, 600 ms, 800 ms, or 3,000 ms. The version administered to each participant was determined at random.

This task measures racial bias by comparing participants' accuracy (i.e. correct responses) on stereotype-incongruent trials (those with a Black prime and a tool or a White prime and a gun) to accuracy on stereotype-congruent trials (those with a Black prime and a gun or a White prime and a tool). An accuracy bias score was calculated for each participant by subtracting the average accuracy on stereotype-incongruent trials from the average accuracy on stereotype-congruent trials. This gave a higher accuracy bias score to participants who made more stereotypical errors on stereotype-incongruent trials. The reliability of the WIT for accuracy bias has been shown to be .74 in previous research (Ito et al., 2013).

Although accuracy served as a single dependent measure, participants' response times were also used to create a second measure of racial bias, reaction time bias. The reaction time bias score was calculated by subtracting the average reaction time on stereotype-congruent trials from stereotype-incongruent trials. This gave a higher reaction time bias to participants who responded faster on stereotype-congruent trials than on stereotype-incongruent ones. The reliability of the WIT for reaction time bias has been shown to be .46 in previous research (at a response window length of 500 ms in Ito et al., 2013).

*First-Person Shooter Task (FPST):* In this task, participants were instructed to make a decision to press a button labeled "shoot" or "don't shoot" based on whether the person in the image on the computer screen was holding a gun or something that was not a gun (e.g. a wallet, cell phone, or soda can), respectively. Participants were told that the goal of the task was to shoot anyone holding a gun – the "bad guys." The program flashed different background scenes (each was shown for 500 to 800 ms) and then inserted a person into the scene holding either a gun or a different object (shown for 590 ms), at which time the participants would press one of the two



buttons on the button box. The participants were given 12 practice trials followed by one block of 96 actual trials.

If the participants responded correctly within the 590 ms response window deadline, they would receive 5 points if the object was not a gun or 10 points if the object was a gun. If the participants responded incorrectly, they would lose 20 points if the object was not a gun or 40 points if the object was a gun. If the participants responded too slowly, they would lose 50 points, and were instructed that because the penalty for being too slow was so high, it was very important for them to respond quickly. A different sound would play each time the participants responded correctly, incorrectly, or too slowly. Current points, the amount of points gained or lost, and a feedback message were shown after every response for 500 ms. The feedback messages were “Good shot” for correct shoot responses, “You shot a good guy!!” for incorrect shoot responses, “Wise choice” for correct don’t shoot responses, and “YOU’RE DEAD!!” for incorrect don’t shoot responses. Total points were also shown at the end of the task.

This task measures racial bias in the same way as the WIT: an accuracy bias score was calculated by subtracting the average accuracy on stereotype-incongruent trials from the average accuracy on stereotype-congruent trials. The higher the accuracy bias scores, the more stereotypical errors the participants made on stereotype-incongruent trials. The reliability for the accuracy bias on the FPST has been shown to be .09 in previous research (Ito et al., 2013). Similar to the accuracy bias, a reaction time bias score was calculated by subtracting average reaction time on stereotype-congruent trials from average reaction time on stereotype-incongruent trials, such that higher reaction time bias scores indicated faster responses to trials consistent with stereotypes. The reliability for reaction time bias on the FPST has been shown to be .07 in previous research (Ito et al., 2013).

Finally, participants completed a number of direct explicit measures to measure their overt attitudes toward Blacks and motivation to control prejudice. The direct explicit tasks included the Motivation to Control Prejudiced Reactions Scale (MCPR; Dunton & Fazio, 1997; reliability of .81 in Ito et al., 2013), the Internal and External Motivation to Respond without Prejudice Scales (iMCPR and eMCPR; Plant & Devine, 1998; reliability of .85 and .83, respectively in Ito et al., 2013), the Attitude Toward Blacks Scale (ATB; Brigham, 1993; reliability of .86 in Ito et al., 2013), and a series of feeling thermometers that asked participants to rate the “warmth” of their feelings toward different social groups (including African Americans and Caucasians). Sample items from each task are included in Appendix A. Participants were also asked for their race and gender.

## Results

First, the effect of the WIT response window length manipulation was examined on the WIT accuracy (ACC) and reaction time (RT) biases, using separate univariate ANOVAs. The WIT response window was the only factor included in these analyses. The main effect of response window on ACC bias was significant,  $F(3, 110) = 5.46, p < .01$  (see Figure 2), indicating that participants had different racial biases on the WIT for the four response window lengths. Participants had higher ACC biases for shorter response window lengths than for longer response window lengths (see Table 1). In order to examine which versions of the task induced a significant amount of racial bias in responses, four separate post-hoc one-sample t-tests were run, one for each condition. Only the 400 ms response window length was significantly different from zero,  $t = 3.08, p = .01$ , indicating that ACC bias was only present in the 400 ms condition (see Table 1).

The main effect of WIT response window on RT bias was also significant, indicating that participants had different racial biases for the different WIT response window lengths,  $F(3, 110) = 2.77, p = .05$ . Participants had higher RT biases for longer response window lengths (see Figure 3). The four separate post-hoc one-sample t-tests showed that only the 800 ms condition was significantly different from zero,  $t = 2.38, p = .03$ , which indicates that RT bias was only present in the 800 ms condition (see Table 1).

Next, how participants' executive function (EF) ability influenced the amount of bias they showed in each WIT condition was examined. This was done using a univariate ANOVA that allowed for a categorical variable (WIT response window) to interact with a continuous variable (EF ability). In order to create a single measure of general EF, the scores from the two EF tasks (Antisaccade and Spatial Span) were converted into z-scores and then averaged. The EF tasks were significantly correlated with one another,  $r = .23, p = .01$ . The main effect of EF on ACC bias was not significant,  $F(1, 105) = 1.67, p = .20$ . The main effect of EF on RT bias was also not significant,  $F(1, 105) = 0.19, p = .67$ . The interaction of EF ability and WIT response window was not significant for ACC bias,  $F(3, 105) = 0.74, p = .53$ , but was significant for RT bias,  $F(3, 105) = 3.11, p = .03$ .

In order to examine this interaction, WIT RT biases for the four different response windows were then compared to EF ability using basic bivariate correlation analyses. The EF composite variable was negatively correlated with RT bias for the 3,000 ms condition,  $r = -.37, p = .05$  (Table 2), which indicates that participants with higher EF ability had lower reaction time biases when there was sufficient time for them to use their executive control. This was the only significant correlation to emerge, but the EF composite correlation with RT bias at 800 ms was marginally significant,  $r = .37, p = .07$ , albeit in the opposite direction of what was expected.

Finally, the relationships between the WIT responses and responses on the different implicit (FPST), indirect explicit (Funding and Seating tasks), and direct explicit tasks (MCPR, iMCPR and eMCPR, ATB, and the feeling thermometers) were examined. No significant correlations were found between the FPST and the WIT (see Table 3). This indicates that, in the current experiment, biases on the FPST were not equivalent to those on the WIT.

The WIT biases were initially compared to the different explicit tasks on their own, only revealing a negative correlation between the Funding task and RT bias in the 600 ms response window,  $r = -.40, p = .03$ . There were also a few marginally significant findings ( $ps \leq .08$ , see Table 4). Then, the WIT biases were compared to the different explicit tasks using two composite variables, one for Indirect Explicit Bias and one for Direct Explicit Bias, to see if any other relationships would emerge. The scores on the different explicit tasks were converted to z-scores and scaled such that a higher score indicated a higher level of explicit bias. The correlations indicated that neither the Indirect Explicit Bias nor the Direct Explicit Bias was significantly correlated with the WIT biases for any condition (Table 4). The lack of correlations with the explicit bias tasks shows that bias on these tasks was not predictive of bias on the WIT in the current study.

## DISCUSSION

The goal of the current research was to determine the effect of executive function (EF) ability and response window length on a specific implicit measure of racial bias, the weapon identification task (WIT). In addition, the relationships between the accuracy and reaction time biases on the WIT and various direct and indirect explicit measures were explored. The biases on the WIT were also compared to those on the first-person shooter task (FPST).

The results replicated previous research findings for the relationships between response window length and bias on the WIT (Payne, 2001; Payne et al., 2002), demonstrating that participants have higher accuracy (ACC) biases at shorter response windows and higher reaction time (RT) biases at longer response windows. In addition, the results did not show a significant main effect of EF ability on ACC bias or an interaction between EF ability and response window length on ACC bias, which did not replicate previous research (Payne, 2005; Payne & Gawronski, 2010). There were also no significant correlations between the WIT biases and the FPST biases, which does not replicate previous research (e.g. Ito et al., 2013).

#### *EF Ability and Racial Bias on the WIT*

New results that have not been demonstrated in previous research emerged for the relationship between EF ability and RT bias. There was no main effect of EF ability on RT bias, but there was a significant interaction between EF ability and WIT response window for RT bias. At shorter response windows, EF ability did not affect RT bias ( $r_s = -.07$  and  $.12$  for 400 ms and 600 ms, respectively), but at longer response windows it did ( $r_s = .37$  and  $-.37$  for 800 ms and 3,000 ms, respectively). In the 3,000 ms condition it was observed that participants with higher EF ability had significantly less RT bias than participants with lower EF ability. One possible reason for this is that participants with lower EF ability may have acted more impulsively when stereotypes were activated, while participants with higher EF ability may have been more likely to quickly override the automatic associations on stereotype-incongruent trials. Further examination of the data demonstrated that there was no significant relationship between reaction times on stereotype-congruent trials and EF ability ( $r = -.02, p = .90$ ) or for stereotype-incongruent trials and EF ability ( $r = -.14, p = .46$ ) for the 3,000 ms condition. Though the relationship was not significant, participants with lower EF ability tended to be slower on

stereotype-incongruent trials than on stereotype-congruent trials, whereas participants with higher EF ability tended to have similar reaction times between the two trial types. The fact that participants with lower EF ability were slowed down by stereotype-incongruent trials suggests that these participants needed to exert more effort to overcome their automatic responses than participants with higher EF ability.

However, in the 800 ms condition, there was a positive correlation between EF ability and RT bias rather than a negative one, which suggests that participants with higher EF ability responded more impulsively than participants with lower EF in this condition. It is unclear why the correlations have opposite signs between the 800 ms and 3,000 ms conditions, but it may be because the 3,000 ms response window length is much longer than any of the other conditions, so there is essentially no response deadline for this condition. Participants with lower EF ability had very similar RT biases across the 400 ms, 600 ms, and 800 ms, but had much higher RT biases for the 3,000 ms condition. The lack of differences between the three shorter response windows suggests that there was not enough time for participants with lower EF to exert executive control and these participants were therefore responding at the same rate to stereotype-congruent and stereotype-incongruent trials. In the 3,000 ms condition, there was sufficient time for lower EF ability participants to exert control, so they made virtually no mistakes. When the participants were able to accurately respond without concentrating on responding within the deadline, they responded faster to stereotype-congruent trials, which led to the higher RT bias that was observed in the 3,000 ms condition for lower EF ability participants.

It was not clear why participants with higher EF ability shifted from gradually increasing RT biases on the 400 ms, 600 ms, and 800 ms conditions to having much less RT bias when the response window was 3,000 ms. Further examination of the data showed that the 3,000 ms

condition did not have any participants with an EF ability larger than one standard deviation above the average, whereas all three of the shorter response window conditions did (at least three in each condition). Therefore, any analyses on how participants with high EF ability performed in the 3,000 ms condition were extrapolated estimates and not indicative of actual participant performance, so no conclusions can be made about how higher EF ability modulates performance in this condition.

Another possibility to consider for the above analyses is that the EF tasks were not good measures of EF ability, and therefore, the analyses using EF ability are not meaningful. This is supported by the low correlation between the Antisaccade and Spatial Span tasks ( $r = .23, p = .01$ ). Previous research (e.g. Miyake et al., 2000) has demonstrated that different EF tasks have highly variable correlations to overall EF ability, especially when they measure different types of EF abilities. Because the current study only included two EF tasks and each task measured different EF abilities (i.e. inhibition and working memory), it is likely that the estimate for EF ability was not accurate. Using more EF tasks that measure similar EF abilities may significantly increase the validity of the results (see Ito et al., 2013).

If EF ability is disregarded, there were still no clear results for the relationship between the different response windows and the ACC and RT biases. Only one of each type of bias was significantly different from zero for any of the response window lengths (400 ms for the ACC bias and 800 ms for the RT bias), which does not replicate previous findings using similar response window lengths – Payne (2001) used a 500 ms response window and an unlimited time condition and Payne et al. (2002) used 200 ms, 450 ms, and 700 ms response window lengths. Both Payne (2001) and Payne et al. (2002) had scores significantly different from zero for all conditions. In the current study, the results only followed the same basic pattern as this previous

work: as the response window length decreased, the RT bias decreased and the ACC bias increased.

The lack of consistency in the WIT biases being significantly different from zero could potentially be due to the low number of participants per condition. But, previous research used approximately the same number of participants per condition with significant results ( $n$  of about 30 participants in both Payne, 2001 and Payne et al., 2002). The standard deviations were smaller for reaction times in Payne (2001), though (about 63.5 ms for the unlimited time condition and 29.3 ms for the 500 ms condition). In the current study, the standard deviation for the 3,000 ms condition was 72.0 ms and averaged 53.0 ms for the 400 ms and 600 ms conditions combined. This indicates that despite the low number of participants, the RT biases in the Payne (2001) study were more consistent. On the other hand, the accuracies in Payne (2001) were comparable to the ones in the current study (accuracy of .09 for the unlimited time condition and .15 in the 500 ms condition in Payne, 2001, compared to an accuracy of .07 for the current study's 3,000 ms condition with an average of .12 for the 400 ms and 600 ms conditions combined). Still, the larger spread of reaction times in the current experiment may have contributed to the lack of consistency in the RT bias results. Having more participants may increase the consistency of RT bias and ACC bias scores (see Ito et al., 2013).

It is also possible that the larger standard deviations for the different conditions are related to exposure to Blacks. The current study was done at the University of Colorado Boulder, while the Payne (2001) and Payne et al. (2002) work were conducted at Washington University in St. Louis, Missouri. Washington University has much more diversity than the University of Colorado, with about 58% Caucasians and 6% African Americans (Washington University in St. Louis, 2010) versus 75% Caucasians and 2% African Americans at the University of Colorado



(College Portrait, 2012). In addition, St. Louis, Missouri has many more African Americans and Blacks – 49.2% (U.S. Census Bureau, 2010b) compared to 0.9% in Boulder (U.S. Census Bureau, 2010a). Because the participants from the current study had much less exposure to African Americans and Blacks, the accessibility of racial stereotypes may have been different for these participants compared to the participants in the Payne (2001) and Payne et al. (2002) studies, which could have impacted all of the results in the current study.

*FPST and Racial Bias on the WIT*

Analyses to determine the relationship of the ACC and RT biases on the FPST with the ACC and RT biases on the WIT demonstrated that none of the FPST biases correlated with the WIT biases for the different response window lengths. This is inconsistent with previous research that has demonstrated a significant, albeit weak, correlation between the two tasks (e.g.  $r = .17, p \leq .05$  in Ito et al., 2013). It is important to note that the low reliability of the FPST ( $r = .09$  for ACC bias and  $.07$  for RT bias, as determined by Ito et al., 2013) may contribute to the lack of relationships found in the current study. Having more participants complete the study would likely yield similar results to previous research.

In addition, previous research has shown that participants can be influenced to respond differently on implicit tasks based on what information is made salient to them (Han, Czeisler, Olson, & Fazio, 2010; see Blair, 2002). Therefore, it is possible that administering the FPST right after the WIT may have led participants to have a different interpretation of the task than if the FPST were to be administered alone. For example, having already completed the WIT may have made participants more likely to think, “mistaking guns for tools makes me look racist when there are Black faces presented before the objects – I should be more careful about shooting Black people on this new task, because I don’t want to look racist.” Or, the WIT may

have simply made race more salient on the FPST. It is unclear which possibility is more likely, because neither of the FPST biases was significantly different from zero. The ACC bias was marginally significant, but negative, which indicates less racial bias on the task,  $t = -1.91$ ,  $p = .06$ . The RT bias was not significant,  $t = 0.17$ ,  $p = .87$ , indicating no racial bias in terms of reaction times on the task. Previous research has demonstrated that making race salient increases bias (e.g. Payne et al., 2002), but the results of the current study do not replicate this finding.

#### *Explicit Measures and Racial Bias on the WIT*

Analyses on the relationship between the explicit tasks and the WIT biases revealed only one significant correlation. The Funding task was negatively correlated with RT bias at the 600 ms response window. Because there were no other significant correlations (although there were a few marginally significant findings), this result was likely a Type I error. The explicit biases were converted into two composite variables (one for Indirect Explicit Bias, the other for Direct Explicit Bias) to see if there were any overall correlations. Neither of these two composite variables correlated to the WIT ACC or RT biases. Previous research has demonstrated that there are weak correlations between implicit bias scores and scores on explicit measures (Payne, 2001; see Payne, Burkley, & Stokes, 2008). The current study sporadically replicated these findings.

Similarly, previous research using latent variable analyses has shown that an internal motivation to control prejudice is negatively correlated with implicit bias task scores and that an external motivation to control prejudice is positively correlated with implicit bias scores (Ito et al., 2013). Latent variable analysis could not be used in the present study due to the low number of participants in each condition. It is possible that more consistent correlations would have emerged if there were a sufficient number of participants to conduct this more advanced analysis.

*Practical Applications of Implicit Bias Tasks*

While the current research did not demonstrate clear relationships between executive control and bias on the WIT, previous research has shown that EF ability can be used to predict ACC bias (Payne, 2005). Therefore, it is important to consider the practical applications of tasks such as the WIT and FPST rather than only looking at their theoretical implications. Classic studies such as Word, Zanna, and Cooper (1974) demonstrate that automatic stereotyping can have a significant impact on how people present themselves and how others view them in return. In the first experiment of the Word et al. (1974) work, White participants in the position of interviewers tended to rate Black confederates in the position of interviewees as less qualified for a job position. The behavior of the participants and the confederates was analyzed, and it was found that the participants sat further away from the Black interviewee (compared to the White interviewee), conducted shorter interviews, and made more speech errors. In the second experiment, White confederates were put in the role of interviewers to see whether there was a difference in interviewee performance when the interviewer sat further away from the interviewee and made more speech errors (as the participants in the first experiment had behaved when interviewing the Black confederates). The participants who were interviewed in this manner were subsequently rated as less qualified, and they reported feeling less calm during the interview and having a more negative mood following the interview. The Word et al. (1974) study demonstrated that people's actions may be unintentionally discriminatory and that these actions may make receivers anxious, which will further decrease people's impressions of them.

Other studies have indicated that implicit racial bias scores are significantly correlated with nonverbal friendliness of interactions with Blacks (Dovidio, Kawakami, & Gaertner, 2002; McConnell & Leibold, 2001). These studies, in conjunction with the Word et al. (1974) and

Devine (1989) work predict that stereotypes can unintentionally influence people's actions toward and evaluations of other people. Unintended prejudice and discrimination can have unfortunate consequences for well-qualified job applicants, patients seeking medical assistance, and others, simply because of their race, gender, or other group membership (see Jost et al., 2009 for a list of ten important studies).

The importance of unintended prejudice and discrimination is especially critical in split-second situations that can mean the difference between life and death, as when police need to make a decision to shoot a person who is perceived to be holding a gun. The media has presented many cases of police mistakenly shooting unarmed African Americans and Blacks, which demonstrates that automatic associations can potentially influence split-second decisions to shoot suspects (Correll et al., 2002; Payne, 2001; Plant & Peruche, 2005). Although it is not possible to directly measure how automatically activated stereotypes influence real-life decisions, it is important to theoretically assess how executive control plays a role in reducing the influence of automatic associations in order to prevent unintended discrimination. Plant and Peruche (2005) found that with practice, police officers will show less racial bias on the FPST, indicating that the criteria officers use to shoot are influenced by their ability to control their responses. The Plant and Peruche (2005) work supports the idea that automatic prejudice and discrimination can be reduced. Still, the FPST is only a theoretical measure of discrimination, and reimplementing the task in a more real-world scenario or simulation might provide a better picture of how EF ability influences split-second decisions.

### *Conclusion*

Although the current study cannot be used to address how unintentional prejudice and discrimination affect people in real-world situations, it is useful on a theoretical level. Each

participant completed two EF tasks (to measure participants' abilities to control their responses) as well as one version of the WIT (as a measure of implicit racial bias). There were few significant results, but many of the patterns found in this study replicate previous research and demonstrate that EF ability plays an important role in modulating how people respond to stereotype-congruent versus stereotype-incongruent information when there is a time constraint. However, further research needs to be done before any conclusions can be drawn from the current study.

Future research should include more participants per condition and more EF tasks, which will increase the possibility of obtaining significant results. It may also be useful to prescreen participants for their motivation to control prejudice before participating in the study. Preselecting participants based on their prejudice levels was used in the Devine (1989) work to obfuscate the purpose of the studies, and a similar procedure could be used with the present methodology to ensure that performing the WIT and FPST would not influence answers on the explicit measures. The above changes would significantly increase the validity of the results and give further insight into how to study implicit racial bias as well as what potential avenues exist for reducing its influence on real-world situations.

## REFERENCES

- Blair, I.V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6(3), 242-261.
- Brigham, J.C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology*, 23, 1933-1967.
- College Portrait. (2012). University of Colorado Boulder College Portrait. Retrieved March 19, 2013, from <http://www.collegeportraits.org/CO/CU-Boulder/characteristics>
- Correll, J., Park, B., Judd, C.M., & Wittenbrink, B. (2002). The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. *Journal of personality and social psychology*, 83(6), 1314-1329.
- Correll, J., Park, B., Judd, C.M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*, 37(6), 1102-1117.
- Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5-18.
- Dovidio, J.F., Kawakami, K., & Gaertner, S.L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82(1), 62-68.
- Dunton, B.C., & Fazio, R.H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 23, 316-326.
- Friedman, N.P., Miyake, A., Young, S.E., DeFries, J.C., Corley, R.P., & Hewitt, J.K. Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology*, 137(2), 201-225.
- Greenwald, A.G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4-27.

- Han, H.A., Czellar, S., Olson, M.A., & Fazio, R.H. (2010). Malleability of attitudes or malleability of the IAT? *Journal of Experimental Social Psychology*, 46, 286-298.
- Ito, T.A., Bartholow, B.D., Correll, J., Friedman, N.P., Miyake, A., Altamirano, L., & Loersch, C. (2013). *Mapping implicit bias: Using a latent variable approach to understand the role of cognitive control in implicit bias*. Manuscript in preparation.
- Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513-541.
- Jost, J.T., Rudman, L.A., Blair, I.V., Carney, D.R., Dasgupta, N., Glaser, J., & Hardin, C.D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, 29, 39-69.
- Klauer, K.C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *The Quarterly Journal of Experimental Psychology*, 63(3), 595-619.
- McConnell, A.R., & Leibold, J.M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37(5), 435-442.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.
- Payne, B.K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81(2), 181-192.

- Payne, B.K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, 89(4), 488-503.
- Payne, B.K., Burkley, M.A., & Stokes, M.B. (2008). Why do implicit and explicit attitudes diverge? The role of structural fit. *Journal of Personality and Social Psychology*, 94(1), 16-31.
- Payne, B.K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B.K. Payne, & B. Gawronski (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (1-14). New York: Guilford Press.
- Payne, B.K., Lambert, A.J., & Jacoby, L.L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology*, 38, 384-396.
- Plant, E.A., & Devine, P.G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811-832.
- Plant, E.A., & Peruche, B.M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science*, 16(3), 180-183.
- Siegel, E.F., Dougherty, M.R., & Huber, D.E. (2012). Manipulating the role of cognitive control while taking the implicit association test. *Journal of Experimental Social Psychology*, 48, 1057-1068.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology*, 125(1), 4-27.



- U.S. Census Bureau. (2010a). *Profile of general population and housing characteristics: Boulder, CO*. Retrieved March 22, 2013, from [http://factfinder2.census.gov/bkmk/table/1.0/en/DEC/10\\_DP/DPDP1/1600000US0807850](http://factfinder2.census.gov/bkmk/table/1.0/en/DEC/10_DP/DPDP1/1600000US0807850)
- U.S. Census Bureau. (2010b). *Profile of general population and housing characteristics: St. Louis, MO*. Retrieved March 22, 2013, from [http://factfinder2.census.gov/bkmk/table/1.0/en/DEC/10\\_DP/DPDP1/0500000US29510](http://factfinder2.census.gov/bkmk/table/1.0/en/DEC/10_DP/DPDP1/0500000US29510)
- Washington University in St. Louis. (2010). Demographics and academic divisions. Retrieved March 19, 2013, from <http://careercenter.wustl.edu/whywustl/Pages/Demographics.aspx>
- Wittenbrink, B., Judd, J.M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81(5), 815-827.
- Word, C.O., Zanna, M.P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10(2), 109-120.

## TABLES

Table 1.

*WIT Accuracy Bias and Reaction Time Bias Descriptive Statistics*

Response		Bias Test			
Window	Bias Type	Mean (SD)	N	T-value	P-value
400 ms	ACC	0.10 (0.17)	29	3.08	<b>.01**</b>
	RT	-4.43 (18.00)	29	-1.33	.20
600 ms	ACC	0.01 (0.10)	30	0.41	.69
	RT	-0.08 (17.37)	30	-0.03	.98
800 ms	ACC	0.02 (0.05)	26	1.93	.07
	RT	7.73 (16.54)	26	2.38	<b>.03*</b>
3000 ms	ACC	< 0.01 (0.04)	29	0.09	.93
	RT	7.87 (24.62)	29	1.72	.10

*Note.* ACC = accuracy, RT = reaction time. The Bias Test determined whether the amount of bias was significantly different from zero (i.e. one-sample t-test).

Significant findings are in bold, \* $p < .05$ , \*\* $p \leq .01$ .

Table 2.

*WIT Bias Correlations to Executive Function Tasks*

Response Window	ACC Correlation	RT Correlation
400 ms	.22	-.07
600 ms	.17	.12
800 ms	.02	<b>.37<sup>†</sup></b>
3000 ms	.01	<b>-.37*</b>

*Note.* ACC = accuracy, RT = reaction time.

Significant and marginally significant findings are in bold,  $*p < .05$ ,  $^{\dagger}p = .07$

Table 3.

*WIT Bias Correlations to FPST Biases*

Response Window	FPST Bias	ACC Correlation	RT Correlation
400 ms	ACC	.03	-.16
	RT	-.33	-.03
600 ms	ACC	.23	-.19
	RT	.05	.08
800 ms	ACC	.31	-.13
	RT	.08	.01
3000 ms	ACC	-.18	.24
	RT	-.23	.13

*Note.* ACC = accuracy, RT = reaction time.

Table 4.

*WIT Bias Correlations to Explicit Tasks*

Response Window	Explicit Variable	ACC Correlation	RT Correlation
400 ms	<u>Direct Composite</u>	.20	-.02
	MCPR	-.14	.01
	eMCPR	-.21	.14
	iMCPR	-.12	.08
	ATB	.20	-.05
	Feeling Thermometers	.08	.18
	<u>Indirect Composite</u>	-.16	-.11
	Funding Task	-.07	-.07
	Seating Task	-.12	-.07
600 ms	<u>Direct Composite</u>	.01	-.27
	MCPR	.04	.15
	eMCPR	.11	.01
	iMCPR	.14	.24
	ATB	.07	<b>-.34<sup>†</sup></b>
	Feeling Thermometers	.24	-.07
	<u>Indirect Composite</u>	.15	-.24
	Funding Task	-.13	<b>-.40*</b>
	Seating Task	<b>.35<sup>†</sup></b>	.07
800 ms	<u>Direct Composite</u>	-.04	-.33
	MCPR	-.16	.02
	eMCPR	-.19	.23
	iMCPR	.16	.26
	ATB	-.13	-.21
	Feeling Thermometers	-.16	-.27
	<u>Indirect Composite</u>	-.21	-.25
	Funding Task	-.18	-.14
	Seating Task	-.14	-.25
3000 ms	<u>Direct Composite</u>	-.29	.22
	MCPR	<b>.33<sup>†</sup></b>	.03
	eMCPR	<b>.35<sup>†</sup></b>	-.13
	iMCPR	.31	-.21
	ATB	-.03	.20
	Feeling Thermometers	.29	.15
	<u>Indirect Composite</u>	.18	-.14
	Funding Task	.09	< -.01
	Seating Task	.16	-.19

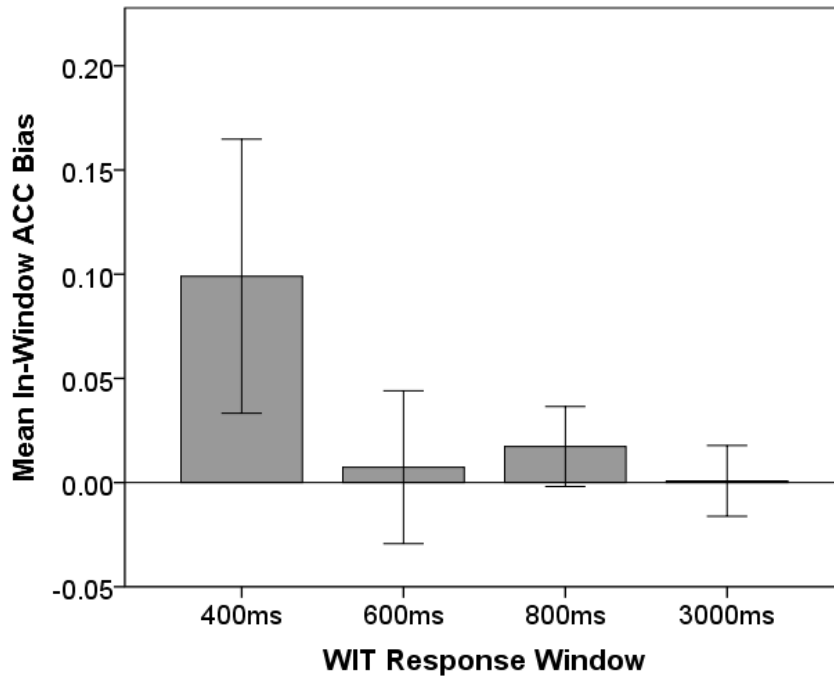
*Note.* ACC = accuracy, RT = reaction time, MCPR = Motivation to Control Prejudiced Reactions Scale, eMCPR = External Motivation to Respond without Prejudice Scale, iMCPR = Internal Motivation to Respond without Prejudice Scale, ATB = Attitude Toward Blacks Scale.

Significant and marginally significant findings are in bold, \* $p < .05$ , <sup>†</sup> $p \leq .08$

FIGURES

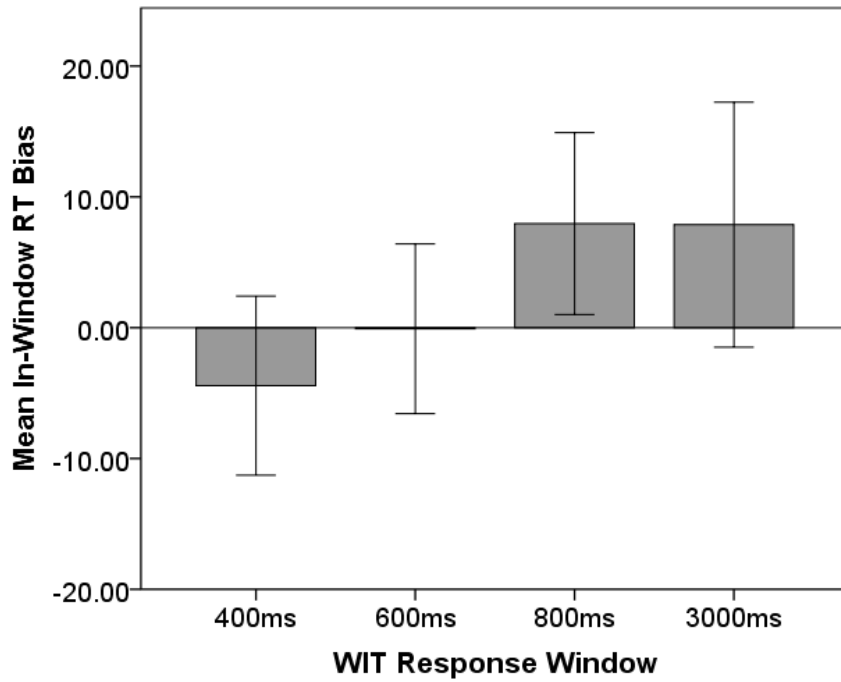


*Figure 1.* Example scene from the Seating task.



*Figure 2.* Mean in-window accuracy bias for all WIT conditions.

ACC = accuracy. Error-bars represent the 95% confidence interval.



*Figure 3.* Mean in-window reaction time bias for all WIT conditions. RT = reaction time. Error-bars represent the 95% confidence interval.



## APPENDIX A

*Motivation to Control Prejudiced Responses Sample* (nine-point Likert scale from strongly disagree to strongly agree):

- “Because of today’s PC (politically correct) standards, I try to appear nonprejudiced toward Black people.”
- “According to my personal values, using stereotypes about Black people is ok.”
- “I am personally motivated by my beliefs to be nonprejudiced toward Black people.”

*Internal and External Motivation to Control Prejudice Sample* (seven-point Likert scale from strongly disagree to strongly agree):

- “I feel guilty when I have a negative thought or feeling about a Black person.” (Internal Motivation)
- “It bothers me a great deal when I think I’ve offended someone, so I’m always careful to consider other people’s feelings.” (Internal Motivation)
- “It’s important to me that other people not think I’m prejudiced.” (External Motivation)
- “I’m not afraid to tell others what I think, even when I know they disagree with me.” (External Motivation)

*Attitude Toward Blacks Scale Sample* (seven-point Likert scale from strongly disagree to strongly agree):

- “I would rather not have Black people live in the same apartment building I live in.”
- “I get very upset when I hear a White person make a prejudicial remark about Black people.”

- “The federal government should take decisive steps to override the injustices Black people suffer at the hands of local authorities.”

*Feeling Thermometer Sample* (one-hundred-point scale from very coolly to very warmly):

- “How warmly or coolly do you feel about... African Americans?”
- “How warmly or coolly do you feel about... Gay men?”
- “How warmly or coolly do you feel about... Hispanics?”